A Stereo Image Processing System for Visually Impaired

G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and Sazali Yaacob

Abstract—This paper presents a review on vision aided systems and proposes an approach for visual rehabilitation using stereo vision technology. The proposed system utilizes stereo vision, image processing methodology and a sonification procedure to support blind navigation. The developed system includes a wearable computer, stereo cameras as vision sensor and stereo earphones, all moulded in a helmet. The image of the scene infront of visually handicapped is captured by the vision sensors. The captured images are processed to enhance the important features in the scene in front, for navigation assistance. The image processing is designed as model of human vision by identifying the obstacles and their depth information. The processed image is mapped on to musical stereo sound for the blind's understanding of the scene infront. The developed method has been tested in the indoor and outdoor environments and the proposed image processing methodology is found to be effective for object identification.

Keywords—Blind navigation, stereo vision, image processing, object preference, music tones.

I. INTRODUCTION

MUCH of the information that humans get from the outside world is obtained through sight. Without this facility, visually impaired people suffer inconveniences in their daily and social life. A total loss of eyesight is one of the most serious misfortunes that can befall a person. In 2000, World Health Organization (WHO) estimated the blind population to be around 50 million with a further 110 million cases of low vision that are at risk of becoming blind [1]. Currently there are about 55 million blind people in the world and this population is estimated to be 75 million by 2020.

Electronic Travel Aids (ETAs) are the devices that aim at

G. Balakrishnan is with the School of engineering and Information Technology, University Malaysia Sabah, 88999, Kotakinabalu, Malaysia. (phone: +6088-320000-3558, fax: +6088-320348, e-mail: g_krishbalu@yahoo.com).
G. Sainarayanan is with the School of engineering and Information Technology, University Malaysia Sabah, 88999, Kotakinabalu, Malaysia. (e-mail: jgksai@ums.edu.my).
R. Nagarajan is with the School of engineering and Information Technology, University Malaysia Sabah, 88999, Kotakinabalu, Malaysia. (e-mail: nagaraja@ums.edu.my).
Sazali Yaacob is with the Northern Malaysia University College of Engineering, 01000 Kangar, Perlis, Malaysia. (e-mail: sazali@kukum.edu.my).

conveying information about the environment to visually impaired individuals, so that they can exploit a part of the information that sighted people normally used to experience and navigate. Many devices exist to assist visually impaired people for navigation. A number of research institutes and software companies are working on solutions to the problems of determining appropriate navigational information for visually impaired people. From the early 1950's several efforts in the provision of travel aids for visually impaired people have been made [2]. They range from the simple cane to advanced computer based aids. While the development of other assisting devices such as Braille electronic reading machine to aid visually impaired people in their everyday life has been very effective. But the adequate solutions for navigation assistance for visually impaired have not yet been achieved.

Early ETAs in the mid-seventies use ultrasonic sensors to detect the obstacles in their path and information is conveyed to the blind user as vibration signals or as sound signals [3]-[5]. Some of the early ETAs are Batcane, Columbia talking compass, Guidecane, Lasercane, Mowat sensor, the Sonicguide and the Polaran [6]. One of the widely used mobility aid among them is Sonic Guide of Leslie Kay [7], in which echolocation employing ultrasonic waves is used to present the user with an audio representation of their surroundings. With this device the user can learn to replace partially their lost sense of sight as they have been shown to have an improved spatial understanding and awareness. The main drawbacks of early ETAs are that these devices often provide little benefit due to various reasons such as restricted mobility, multiple ultrasonic reflections, commercially expensive and cosmetically unattractive. Some of the ETAs transfer the feedback signal to the blind in terms of vibrations, but they did not prove to be effective for long run. Since sound is less sensible to occlusion, and it allows the simultaneous transmission of multiple streams that can be easily segregated or aggregated by humans, the auditory feedback is effective compared to that of tactile feedback.

With the advancement of digital world, the research efforts are being directed to produce improved navigation aid systems, in which digital video cameras are used as vision sensors. Environmental sensing with a camera is attractive due to wealth of information available through this sensor, its closeness in function to the human eye, typical low power consumption and its low cost.

II. Early Work with Vision Sensors

The concept of using video camera as vision sensor had been introduced in Peter Meijer's portable system, The vOICe [8]. Here the image is captured using single video camera mounted on a headgear and the captured image is scanned from left to right direction for sound generation. The top portion of the image is converted into high frequency tones and the bottom portion into low frequency tones. The loudness of sound depends on the brightness of the pixel.

In this work, the pixel intensity is directly converted into loudness without any image processing effort to enhance the object properties. The sound produced from the unprocessed image will contain more information of the background rather than object. Thus, it will be difficult for the blind user to assimilate the information needed for navigation.

Similar work has been carried out in Navigational Assistance for Visually Impaired (NAVI) [9] where the captured image is resized to 32 X 32 and the gray scale of the image is reduced to 4 levels. With the help of image processing techniques the image is differentiated into objects and background. The processed image is converted into stereo sound where the amplitude of the stereo sound is made directly proportional to intensity of image pixels, and the frequency is inversely proportional to vertical orientation of pixels. Both in The vOICe and NAVI the distance between the user and the obstacle cannot be obtained directly by the users.

A. Stereo Vision Systems

The distance is one of the important aspects for collision free navigation for blinds. In order to incorporate the distance information, stereo cameras are used. Use of stereo vision sensors has been of research interest mainly for extraction of 3D models of objects and for perception of depth. The manner in which human beings use their two eyes to see and perceive the three-dimensional world has inspired the use of two cameras to model the world in three dimensions. The different perspectives of the same view seen by two cameras lead to a relative displacement of the same objects or the same points in world reference (called disparity). The size and direction of these disparities can be utilized for depth estimation [10]. The depth of a point is inversely proportional to the disparity.

Only limited research has been done in blind navigation using stereo cameras. In Optophone [11], to obtain a depth map an edge detection routine is applied to images from two cameras. Disparity is calculated using the edge features of both the images. The depth map is then converted into sound using the method applied in The vOICe system. In The vOICe system, the disparity map of all the edge features in the images is obtained. The user will have difficulties to locate the object since unwanted edge features will also exist. The information about the edges is also not sufficient for detection of objects.

Another pioneering work by Zelek et.al. involves stereo cameras and was designed to provide information about the environment through tactile feedback to the blind [12]. The system comprises of a laptop, a stereo setup with two cameras and a virtual touch tactile system. The tactile system is made up of piezo-electric buzzers attached to each finger on a glove worn by the user. Here the cameras capture images, and the disparity is calculated from those images. The depth information is conveyed to the user by stimulating the fingers. In this work, the location of obstacles are not given importance and the system suffers in stereo matching due to improper balance of stereo cameras.

Another important work reported in this area is the visual support system developed by Yoshihiro Kawai and Fumiaki Tomita [13]. The prototype system has a computer, stereo camera system with three small cameras, headset with a microphone and a headphone and a sound space processor. The images captured by the cameras are analyzed to obtain 3D structure, and object recognition is performed. The results are then converted to user via 3D virtual sound. Here, the image processing method is performed after depth calculation, which results in increased computation time. Also the prototype developed is not portable and can be applicable only in indoor environment with restrictions.

III. Image Processing in Blind Navigation

The fundamental purpose of blind navigation is to assist the blind people to navigate freely among obstacles by providing them the position, size and distance of the obstalces. The obstacles in the captured image should be given more importance compared to that of background. Thus, it is essential to develop a navigation kit such that the blind user understands the environment infront of him with minimum efforts.

Image processing plays an important role in blind navigation. From literature it is evident that the obstacles are not given preference due to the lack of image processing methodologies in blind navigation applications. It is difficult to apply image processing methods in blind navigation due to the varying nature of the environment. The real world scene has to be processed and presented in a more visually recognizable form such as image with only obstacles. It will be sufficient for mobility purposes to have the outline and basic shapes of objects. With the reduction of visual information, only the optimal environmental information which is necessary for navigation is made available to the user.

Since the processing has to be done in real time, the time factor has to be critically considered. Rate of capture of image frames depends on duration of image acquisition, image transfer to computer, image processing, sound generation, sonic transfer to stereo earphones and the duration of acoustic information heard through earphones for each image. However, this rate depends mostly on the time duration of acoustic information sent through the earphones for each image. Proper design of image processing also helps in reducing this computation time.

IV. Experimental Setup

The components used in this work are small enough to be

carried out easily. The system consists of a helmet molded with stereo cameras, wearable computer and stereo earphones. The stereo camera selected is a compact, low-power digital stereo head with an IEEE 1394 digital interface. It consists of two 1.3 mega pixel, progressive scan CMOS imagers mounted in a rigid body, and a 1394 peripheral interface module, joined in an integral unit. The wearable computer [14] used is a compact, high performance 500MHz processor with 256 MB RAM. The helmet can be worn over the head.

The stereo camera is placed in the front and the wearable computer is placed over the top of the helmet as shown in Fig. 1(a). The stereo cameras capture the visual information infront of the blind user. The captured images are then processed using the proposed methodology in the wearable computer. The information about the obstacle is conveyed to the blind user by musical tones through stereo earphones. Fig. 1(b) shows a volunteer wearing the SVETA system.

V. PROPOSED IMAGE PROCESSING

In this paper objects or obstacles are identified and are assigned preference based on their position and distance. The first step in this methodology is the image acquisition, in which scene infront of blind is captured using both cameras simultaneously. The image captured from camera is color image of size 352 x 288. Fig. 2 shows the original left and right images captured from stereo cameras.

Processing the image with original size will increase computation time. Blind navigation requires real time processing. The time factor is very critical in this application. Therefore, pre-processing is undertaken to reduce the computation time, where both the left and right stereo images are converted to gray scale intensity image and resized to 64 x 64. Fig. 3 shows initial processed left and right image of real life picture with three objects.

The main task is to identify and assign importance to the objects based on its distance. In this work, the objects in both the images are identified by locating their boundaries.

1: Wearable computer
2: Stereo camera
3: Helmet
4: Stereo earphone

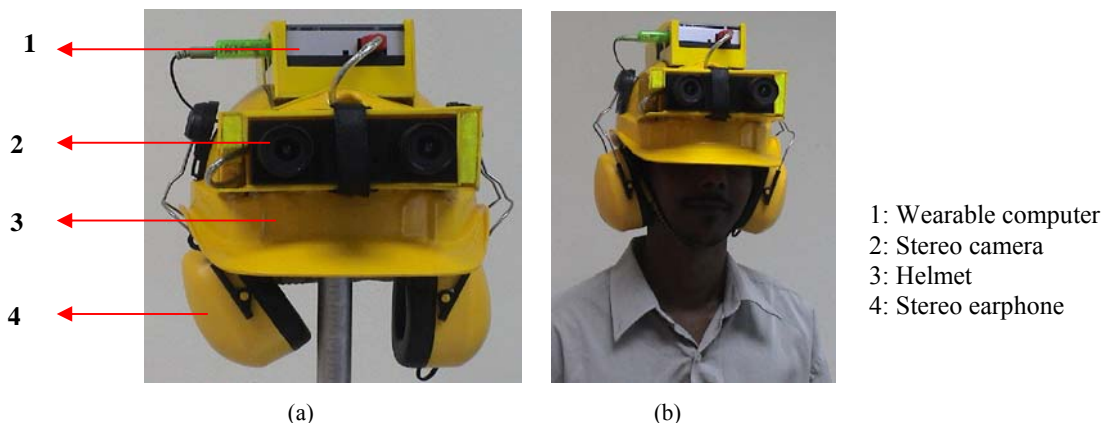(a)   (b)

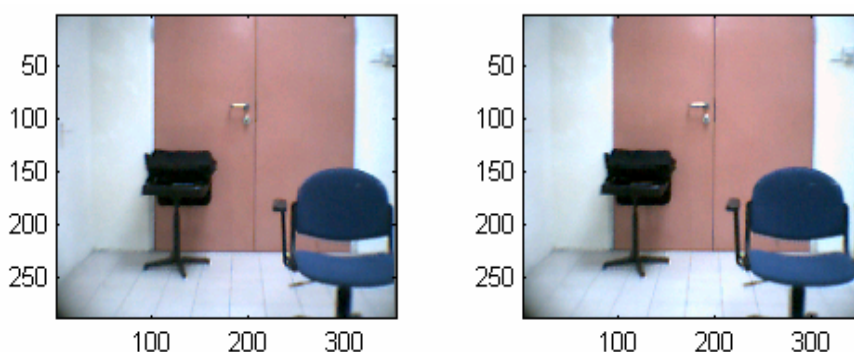Fig. 1 (a) The experimental system. (b) Person wearing the SVETA system
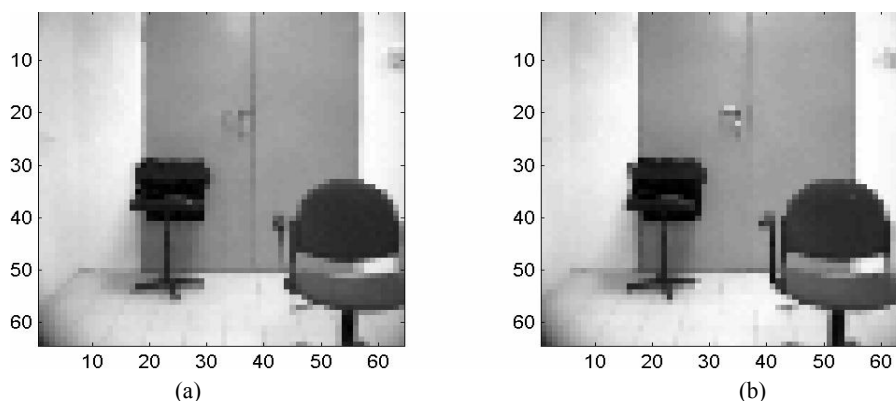

Fig. 2 Original Left and Right Image

Fig. 3 Resized Gray Scale Intensity Image (a) left image (b) right image

A. Edge Detection

Edge detection is one of the important human vision properties. Often human identifies the object with its boundaries and shape. The goal of edge extraction is to provide useful structural information about object boundaries. From the edges, object properties such as area, perimeter and shape can be measured for object recognition.

Canny edge detector is one of the optimum edge detecting method for step edges corrupted by white noise. Canny's method [15] finds edges by looking for local maxima of the gradient of the image. The gradient is calculated using the derivative of a Gaussian filter. This method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise and more likely to detect true weak edges. Thus canny edge detector has better performance in edge detection than the other methods. The edge image of the right camera image using Canny edge detector is shown in Fig. 4.

Fig. 4 Edge features of right camera image

B. Edge Linking

In this work a region within closed edge is considered as an object. It is realized that edge detection alone is not enough to extract the closed boundary of an object from the image. Edge detectors applied to resized grayscale image do not provide enough information to identify meaningful object, since the edges derived can be incomplete and not continuous. Edges in captured images seldom form the closed, connected boundaries that are required for object extraction. In this work, further processing using dilation is undertaken to connect the broken edges into a meaningful object.

Morphological operation using dilation is proposed to link the broken edges. By experimentation it is found that edges of the same objects are usually broken by a maximum of 2 pixels. So a horizontal and vertical disk-structuring element with the size of two pixels is created. This structuring element is used in dilation operation to connect the broken edges around the boundary of all objects. Therefore, the edges around the object that are broken by 2 pixels in horizontal and vertical directions are connected. Fig. 5 shows the image of Fig. 4 obtained after applying the edge linking method.

Fig. 5 Image after edge linking

C. Boundary Handling

In experimentation, it is also noted that the edge detector fails to track edges at the border of image. In vision substitutive system, all objects, regardless of their positions, must be considered. Thus, broken edges at the border of an image should also be connected to form closed objects. All edges in the image are labeled to ensure that only edges with the same label are connected. Once the edges are labeled, each edge in the image border matrix is scanned. Starting at the first scanned edge, the next edges is located and identified as the end edge if it has the same label with the first edge. The pixels between the two scanned edges are set to the intensity similar to that of scanned one. If more than two edges with same label are present at the border, then the process is omitted.

D. Noise Removal

The region inside the closed boundary is considered as an object. The intensity of pixels in the region is enhanced to higher level using flood fill operation. Each object in the image is labeled. It is found that some of the extraneous edges still exist after flood fill operation. The goal of noise removal is to remove extraneous edges present in the image without eliminating the desired objects. Erosion and dilation operations are undertaken to smooth the objects.

A disk-structuring element is selected in such a way that it should be large enough to remove the noise when eroding the image, but not too large to remove the objects. By experimentation a structuring element with the size of three pixels is created. This structuring element is used in erosion operation to remove three pixels from around the boundary of all objects. As a result, extraneous edges presents in the image will be eliminated and the objects will be shrinking. To restore the objects to their original size, dilation operation is applied to the eroded objects using the same structuring element. Thus image with only objects will be obtained. Fig. 6 shows the image obtained after flood fill and noise removal method.

Fig. 6 Binary image after noise removal

VI. ISOLATED OBJECT IMAGE EXTRACTION

Binary images from both left and right cameras with only objects are derived from the previous section. These binary images are mapped with the resized gray scale images to derive new gray scale intensity images with only objects. These processes are preformed for both left and right camera images to obtain two gray scale images with only objects. Fig. 7 shows the left and right gray scale intensity images with only objects.
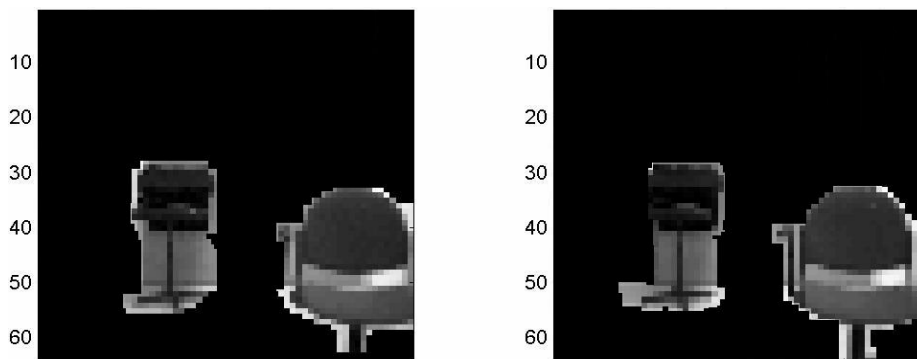
VII. DEPTH CALCULATION

After obtaining the isolated object image, the disparity has to be determined in order to calculate the distance between the blind user and the obstacle. For this, the concept of stereo vision is employed. Stereopsis or stereo vision refers to the problem of determining 3-dimensional structure of a scene from two or more images taken from distinct viewpoints. Given two camera images, if it is possible to identify the image locations that correspond to the same physical point in space from both the camera images, then it is possible to determine its three-dimensional location [10].

The concept of stereo vision is illustrated in Fig. 8. Assume that a point P on a surface is projected on two cameras image planes as P_L (x_l,y_l) and P_R (x_r,y_r) respectively. Let O_L is the optical center of left camera, O_R is optical center of right camera. Here both camera coordinates axes are aligned, and the baseline (line segment joining the optical centers of two cameras) is parallel to the camera x coordinate axis.
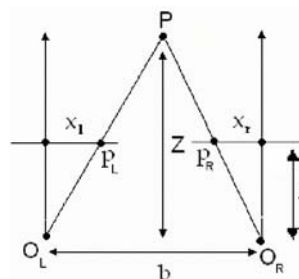

Fig. 8 Stereo Geometry

Given the baseline T (distance between O_L and O_R), and the focal length f of the cameras, depth at a given point may be computed by similar triangles as

$$Z = f\frac{b}{d} \tag{1}$$

where Z is the depth of point P and d is the disparity of P, d=x_l − x_r [16].

Fig. 7 Left and right intensity image with only objects

A. Disparity Calculation

Disparity forms the basic criteria for stereo vision. Once the disparity is determined then the depth can be calculated using equation 1. In order to calculate the disparity, correspondence or matching of two images has to be undertaken. Area based stereo matching is performed over the stereo image pair. Since the cameras are adjusted to same focus by experimentation, the search for correspondence is reduced to a 1-D search. Stereo correspondence is established using SSD (Sum of Squared Differences) correlation method [17]. For each left image pixel, its correlation with a right image pixel is determined by using a correlation window of size 5*5 pixels in which SSD of pixel intensities is computed as

$$c(d) = \sum_{k=-w}^{w} \sum_{l=-w}^{w} \Psi(I_l(i+k, j+l), I_r(i+(k-d_1), j+(l-d_2))) \quad (2)$$

where, *(2w+1)* is the width of the correlation window, I_l and I_r are the intensities of the left and right image pixels respectively,

[i, j] are the coordinates of the left image pixel and
d = [d_1, d_2] is the relative displacement between the left and right image pixels.

Correlation function using Sum of Squared Difference (SSD) is given by

$$\psi(u,v) = -(u-v)^2 \quad (3)$$

Here, the search is reduced to 1-D search. Therefore, d_2 is always equal to 0. The correlation-matching algorithm consists of calculating the correlation values pixel by pixel by varying d_1, which slides the correlation window from the left to the right in the right image along this row. The corresponding point is determined by the pixel that has the highest correlation value. The disparity between the two corresponding points is the distance d_1 that separates them.

To decrease the total computation time, the maximum disparity is bounded to a certain range and thus avoids examining an entire row for each pixel. Since only the isolated object images are used for correspondence, the mismatch error is limited and also the computation time is reduced when compared to conventional area based and feature based techniques. In the disparity map, minimal variation of disparity value if any is noticed. To have an uniform disparity value for an object, histogram is used, where the disparity value that occurs most within the object is found and that value is assigned to all pixels within the object. Fig. 9 shows the disparity map obtained after assigning uniform value. Here the object which is close to the user have high disparity (white color) compared to other objects.
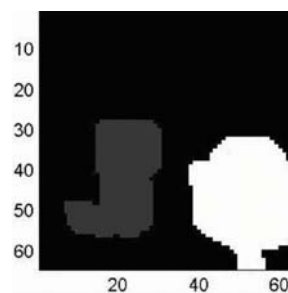

Fig. 9 Disparity Map

VIII. OBJECT PREFERENCE

Real time environment contains more than one object. If all objects are given same preference, then the blind user finds it difficult to identify those objects that are obstacles for navigation. So each object has to be assigned some priority for collision free navigation. In this work objects are given preference based on two important aspects. The first is the distance of the object from the user and the other is the relative position of the object. The object in front and close to the blind user is very important rather than the other objects lying by side and far away from the blind user. In human vision system, the eyes mostly concentrate on a particular object while the background gets less focus. The object of interest is usually at the center of the sight. In this paper, the center of the image is considered as the center of sight. Any object that is located in the central region with high disparity[1] is considered for high preference while the objects located outside the central region with low disparity are less preferred.

In order to locate the object position, object characteristics such as size and centroid are determined. Based on these characteristics, the portion of the object area lying within the central region is calculated. But with these inputs, devising an algorithm to compute object preference is not direct and complex. In order to overcome this uncertainty, fuzzy logic is applied [18]. Four main characteristics are measured for object preference assignment. They are size of an object, Euclidean distance between the object centroid and image centroid, ratio of objects lying within the center of an image and the distance of an object. These characteristics are applied as inputs to fuzzy logic algorithm. Each characteristic is expressed using three membership functions namely low, medium and high. Membership functions are expressed using trapezoidal curve. The output is object preference, which has four trapezoid membership functions such as least preference, medium preference, high preference, and very high preference.

The following table shows the inputs and output of fuzzy rule base. The defuzzification is performed using centroid method. The four inputs generate 81 rules.

---

[1] Disparity is inversely proportional to distance.

TABLE I
INPUTS AND OUTPUT OF FUZZY BASED OBJECT PREFERENCE

Fuzzy Inference System		Membership functions
Input	1. Object size (**OS**)	Low Medium High
	2. Euclidean distance (**ED**)	Low Medium High
	3. Ratio of object area lying within the central region (**RO**)	Low Medium High
	4. Disparity (**DT**)	Low Medium High
Output	1. Object preferences (**OP**)	Least preference Medium preference High preference Very High preference

Some of the fuzzy rules are given below as examples:

Rule 1: If OS is low and ED is low and RO is low and DT is low, then OP is Medium preference.

If object size is small, lying close to the center of image area and has a less disparity value, then it is assigned medium preference.

Rule 6: If OS is low and ED is low and RO is medium and DT is high, then OP is Very High preference.

If object size is small, lying close to the center of image area and has a high disparity value, then it is assigned with very high preference.

Rule 25: If OS is low and ED is high and RO is high and DT is low, then OP is least preference.

If object size is small, lying far from the center of image area and has a less disparity value, then it is assigned with with least preference.

From the fuzzy output the objects are given different intensity values based on their preference. The object with very high preference is symbolized by white intensity and the object with least preference is symbolized with very dark gray intensity. The object preference assigned based on fuzzy rule based system for image of Fig. 9 is shown in Fig. 10. Here, the object in the right has high disparity value and lies in a medium distance from center of image area and hence it is assigned high preference. The other object lies in a medium distance from center of image and has low disparity value, so it is assigned with medium preference.
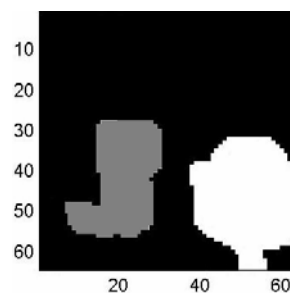

Fig. 10 Image after object preference

Figs. 11, 12 and 13 show the object preference assignment for certain stereo pair images taken in indoor environment. In Fig. 11, the object in the center lies to the center of the image area and has a high disparity value and hence it is assigned with very high preference. The other object lies far from the center of image and has medium disparity value, so it is assigned with medium preference. In Fig. 12, the object in the bottom left has very high disparity value, so it is assigned very high preference even if it doesn't lie to the center of the image area. In Fig. 13, the object in the middle is close to the image area and has high disparity value, so it has high preference compared to other objects.

The fuzzy output image is then converted to musical tones using the sonification method and informed to the blind user through stereo earphone. The sonification method is explained in the following section.
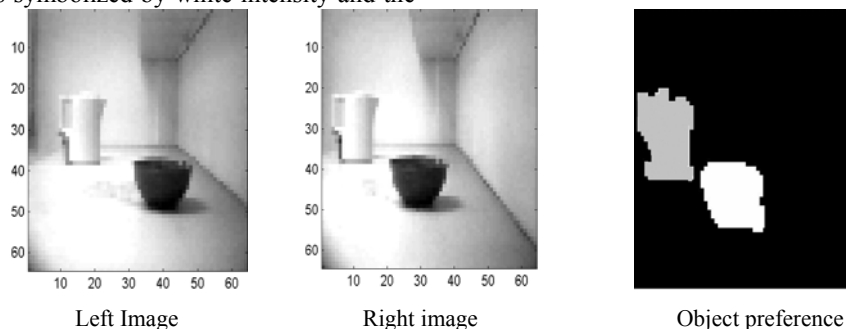

Left Image	Right image	Object preference

Fig. 11 Stereo Pair1 and its object preference

Left Image   Right image   Object preference

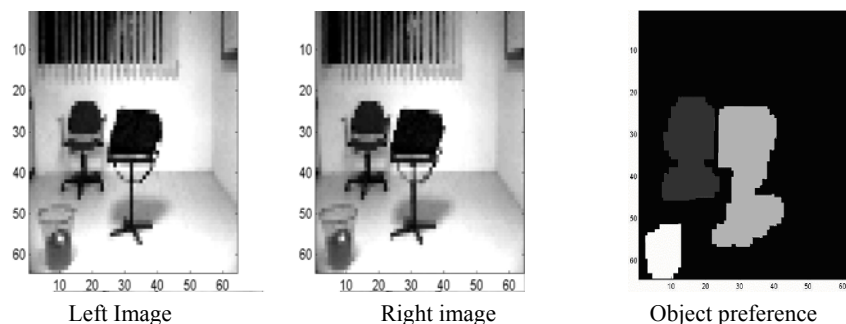Fig. 12 Stereo Pair2 and its object preference

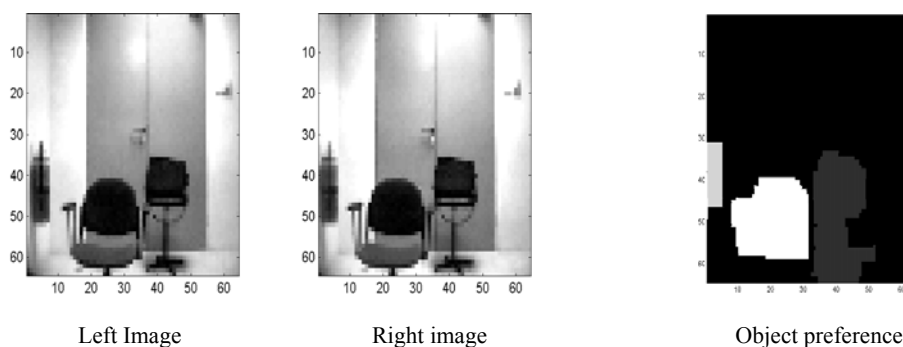Left Image   Right image   Object preference

Fig. 13 Stereo Pair3 and its object preference

IX. MUSICAL TONE BASICS

The sound has several features such as pitch, intensity, timbre and duration. The pitch is the frequency of the sound vibration represented in hertz or cycles. The musical term for frequency is tone. The audible frequency range extends from about 20 Hz to around eight or ten thousand hertz, although this depends entirely on the individual [19]. The audible sensitivity drops with the increase in age. The intensity is the same as loudness and it is related to the amplitude of the sound wave. The other attribute of sound is the duration. Duration is the time during which the specific frequency or tone lasts. Timbre is signature of the source of the sound.

The audible range is divided into octaves. An octave is really a frequency range from a frequency f1 to f2 such that f2 is twice that of f1 in terms of cycles or hertz. The human hearing is logarithmic and is sensitive to frequency octaves. The audible frequency is then comprised of many octaves. Even a frequency range from 20 Hz to 40 Hz is defined as an octave [20]. A piano or a keyboard is a typical western musical instrument. The keys from the left of the keyboard to the right produce higher and higher frequencies. In fact, the key frequencies are arranged in such a manner that they are in a geometric series. That is, the frequency between any key and the key immediately to its left is a constant, the constant being equal to the twelfth root of two or 1.059. For example, typically, there is a white key in the keyboard set to 240 Hz. Then the adjacent key on the right, a black one, is set to 240

X 1.059 = 254 Hertz.

By the specific choice of this ratio (twelfth root of two), the thirteenth key will have the frequency twice that of first key and thus will span a whole octave. In keyboard, the key pattern repeats every twelve keys. This division of the octave into twelve tones is specific to western music. This geometric arrangement of frequencies of the keys in an octave is called an equally tempered arrangement. Most western musical instruments are also tuned to such an arrangement [21].

Even though there is a degree of freedom for selecting the range of an octave (whether it is from 240 to 480 Hz or 254 to 508 Hz etc.), the western music defines a standard octave called the Middle A octave starting from the white key set to 440 Hz. The entire octave (the twelve key patterns) is shown in Fig. 14.
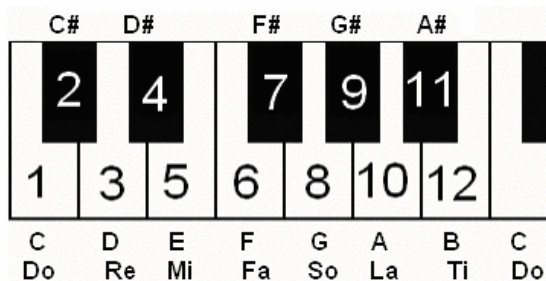

Fig. 14 12 key pattern in piano

From Fig. 14, the keys in the octave have labels for identification. There are seven white keys in an octave, the first one is called C. It progress alphabetically to G and then

back to A and B, after which, the present octave ends and the C key of the next octave begins. The remaining 5 keys are black called C#, D#, F#, G# and A#. The same labeling system is repeated for the keys in the other octaves as well.

Each key is said to be a semitone or half tone apart from its adjacent key. Thus, keys that are second nearest neighbors are considered a whole tone apart. For example, the first white key (C key) and the first black key (C #) are a semitone apart, whereas the first white key (C key) and the second white key (D key) are a full tone apart. The keys and the tones they produce are the basic building blocks of music.

X. PROPOSED SONIFICATION METHOD

With the help of the above rules, musical tones can be incorporated for image sonification. By experimentation it is found that octave frequency of 440 Hz to 880 Hz produces pleasing music. Hence in the developed method, octave frequency of 440 Hz to 880 Hz is selected. With this octave, 12 musical notes are developed. Let f (1,..,12) be the 12 octave frequencies.

The music pattern generated is given by

$$M(j) = \sin(2\pi f(j) t) \qquad j=1,2,\ldots,12 \qquad (4)$$

where

'$M(j)$' is the musical note generated for $f(j)^{th}$ frequency and 't' varies from 0 to desired total duration of the acoustic information presented to the blind.

Different musical tones are generated with the combination of these notes. In this work three notes are combined to form musical tones. Four half steps between first and second note and three half steps between second and third note define major chords. Here eight tones including some major chords are generated using these notes. The image to be sonified is resized to 32 X 32 for reducing the computation time and acoustic duration. Every preceding four rows are grouped and assigned with one musical tone. These musical tones are assigned in such a way that high frequency tones occupy the top portion of the image and low frequency tones are assigned to lower portion of the image. So each pixel in an image is assigned with samples of musical tones based on their position in the image.

The conversion of image into sound involves taking one column at a time starting from left most one and generating sound pattern for that column. The sound pattern generated is given by

$$S(i) = \sum_{j=1}^{32} I(i,j) M(i,j) \qquad i = 1,2,..,32 \qquad (5)$$

where

S(i) is the sound pattern for column i of the image
I(i,j) is the intensity value of (i,j)th element
M(i,j) is the sample of musical tone for (i,j)th pixel.

The sound pattern from each column is appended to construct the sound for the entire image. The scanning of picture is performed in such a way that stereo sound is produced. In this stereo type scanning, the sound patterns created from the left half side of the image is given to left earphone and sound patterns of right half side to right earphone simultaneously. The scanning is performed from leftmost column towards the centre and from right most columns towards the centre, simultaneously. Sound pattern to the left earphone is S_L= S(1) to S(n/2) appended from the left side and sound pattern to the right earphone is S_R= S(n) to S(n/2) appended from the right side, where, n is total number of columns. In this work, n=32.

The sound pattern generated by this sonification method is able to differentiate objects based on its position, shape and distance. The most advantage of this method is that since musical tones are used, the sound generated will be pleasing to the user, less training is needed and continuous use will not fashion loss of interest.

XI. TESTING

For a blind user to navigate freely, the information about the obstacles such as shape, size, position and distance are to be known. If all those information are conveyed to the blind user, then he can move autonomously collision free among obstacles. So, these features are very important for blind navigation. The proposed sonification method is thus compared and tested for their capabilities. In this experimental study, the task of the blind and non blind people was to find obstacles and its characteristics by using the developed sonification methods. The main purpose was to determine whether the subjects can identify the obstacles easily. The experiment was tested using several groups, with 12 people in each group. The data such as object position, size and distance from each method were collected from each subject. Also, every subject is asked a set of questions regarding pleasantness of sound and their response is monitored.

Simulated test images were initially used. They were of white, light gray or dark gray shapes of squares and circles with a black background. White color corresponds to the object in close distance. Different gray shades are assigned to objects based on their corresponding distance. Subjects were also tested with objects of different shapes and sizes placed at different positions, a number of times in different days. The test results obtained are tabulated as in Table II.

TABLE II
TEST RESULTS FOR IDENTIFYING THE CHARACTERISTICS OF OBSTACLES

Object Characteristics	Percentage predicted
	Musical Octave Method
Position (Top/Bottom)	78%
Position (Left/Right)	100%
Shape	88%
Size	91%
Distance	98%
Pleasantness	91%

XII. CONCLUSIONS

In this paper, a scheme for real time image processing and an acoustic transform applied in navigation aid for visually impaired are proposed. It is found that image processing plays an important role in obstacle detection for the navigation of blind. But devising a real time image processing algorithm is relatively complex than conventional offline image processing. In this paper, efforts have been made to incorporate image processing method to develop an effective way for blind navigation. Sonification using musical tones also assists in fast perception of the environment in front of blind user. The proposed methodology runs with a computational time of 1.5 to 2 seconds. Experimentations were conducted with blind volunteers and the blind people were able to navigate in indoor environments and in some of outdoor environments. The works are continued to overcome the problems encountered due to occlusion and improper stereo matching. The work is also being extended in building the system applicable in any environments and much more user friendly.

REFERENCES

[1] World Health Organization (2004) Available: http://www.who.int//mipfiles/2400/AllenFoster.pdf

[2] F. Wong, R. Nagarajan, S. Yaacob, A. Chekima and N. E. Belkhamza, "Electronic *travel aids for visually impaired – A guided tour*", Proceedings of Conference in Engineering in Sarawak, Malaysia, pp. 377-382, 2000.

[3] R. M. Fish, "*Auditory display for the blind*", US Patent No. 3800082, 1974.

[4] C. Capelle, C. Trullemans, "*A Real-Time Experimental Prototype for Enhancement of Vision rehabilitation Using Auditory Substitution*", IEEE Trans. on Biomedical Engineering, Vol. 45, pp. 1279-1293, 1998.

[5] J. Borenstein and I. Ulrich, "*The GuideCane – A computerized travel aid for the active guidance of blind pedestrians*", Proceedings of IEEE International Conference on Robotics and Automation, Albuquerque, NM, pp. 1283-1288, 1997.

[6] S. Shraga, I. Ulrich and J. Borenstein, "*Computerized Obstacle Avoidance Systems for the Blind and Visually Impaired*", Intelligent Systems and Technologies in Rehabilitation Engineering. CRC Press, pp. 414-448, 2000.

[7] L. Kay, "*Electronic aids for blind person: an interdisciplinary subject*", IEE proceeding, Vol. 7, pp. 559-576, 1984.

[8] P. B. L. Meijer, "*An Experimental System for Auditory Image Representations*", IEEE Transactions on Biomedical Engineering, Vol. 39, pp. 112-121, 1991.

[9] G. Sainarayanan, "*On Intelligent Image Processing Methodologies Applied to Navigation Assistance for Visually Impaired*", Ph. D Thesis, University Malaysia Sabah, 2002.

[10] M. Z. Brown and G. D. Hager, "*Advances in computational stereo*", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, pp. 993-1008, 2003.

[11] M. Capp and P. Picton, "*The Optophone: an electronic blind aid*", Engineering Science and education Journal, Vol. 9, No. 3, pp. 137-143, 2000.

[12] J. S. Zelek, S. P. Bromley, D. Aamar and D. Thompson, "*A haptic glove as a tactile vision sensory substitution for way finding*", Journal of Visual Impairment and Blindness, pp 621–632, 2003.

[13] Y. Kawai and F. Tomita, "*A support system for visually impaired persons to understand three-dimensional visual information using acoustic interface*", IEEE conference on Pattern Recognition, Vol. 3, pp. 974-977, 2002.

[14] Xybernaut Wearable computer (2004), Available: http://www.xybernaut.com

[15] J. Canny, "*A Computational Approach to Edge Detection*", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 6, pp. 721-741, 1984.

[16] U. R. Dhond and J. K. Aggarwal, "*Structure from stereo*", IEEE Trans. on systems, man and cybernatics, vol. 19, pp. 1489 – 1510, 1989.

[17] D. A. Forsyth and J. Ponce, "Computer Vision: A modern approach", *Prentice-Hall*, New Jersey, 2002.

[18] T. J. Ross, "Fuzzy logic with engineering applications", *McGraw-Hill*, NewYork, 1997.

[19] R. Elliott, "Fundamentals of music", *Prentice-Hall*, New Jersey, 1971.

[20] Twelve tone technique (2004), Available: http://en.wikipedia.org/wiki/Twelve_tone_technique

[21] J. R. Pierce, "The Science of Musical Sound", *W.H. Freeman & Company*, Newyork, 1992.

G. Balakrishnan is currently a research officer in electrical and electronics engineering in Universiti Malaysia Sabah. He is pursuing his PhD in the same department. He received his B.E., and M.E., degrees, from Baradidhasan University, India, and Bharathiar University, India, respectively in 2001, and 2003. His research interests are in the areas of vision rehabilitation, image processing and machine vision.

G. Sainarayanan is currently a Lecturer in electrical and electronics engineering in Universiti Malaysia Sabah, Malaysia. He received his B.E., M.E., and Ph.D. degrees, respectively, from Annamali University, India, Bharathiar University, India, and University Malaysia Sabah, Malaysia, in 1998, 2000, and 2002. His research interests are in the areas of vision rehabilitation, medical imaging, and intelligent control. He has published more than 45 papers in refereed journals, book chapters, conferences, and research reports.

R. Nagarajan received his B.E. degree (Hons) in electrical and electronics engineering from Madras University in 1961, M. Tech. degree in control engineering from IIT Kanpur in 1969, and Ph.D. degree in adaptive control from Madras University in 1977. Currently he is a Professor in the School of Engineering and Information Technology, University Malaysia Sabah, Malaysia. He was a Visiting Professor at Osaka Prefecture University, Osaka, under the JSPS program from October to November 2004. His current research interests include robot control, stereo image processing, and intelligent process control. He has received several awards for his research contributions which include medals from International Invention and Exhibition in Malaysia and Geneva. He is a Life Fellow of Institution Engineers, India, a Senior Member in IEEE, USA, and a Member of UICEE, Australia.

Sazali Yaacob was born in 1960.He received his Bachelor of Engineering degree in electrical engineering in 1986 from University Malaya. He received his Master of Science degree in system engineering from the University of Surrey, U.K., and later his Doctor of Philosophy degree in control engineering from the University of Sheffield, U.K., in 1987 and 1995, respectively. He was promoted to Associate Professor in 1998 at University Malaysia Sabah, Malaysia, and was consequently appointed as the First Dean of the School of Engineering and Information Technology until 2004. Now, he is working at the Northern University College of Engineering, Malaysia, as Professor in the Center of Engineering Study in Mechatronics. He has publications in the national and international journals and conference proceedings. He has supervised a number of postgraduate students in either Master or Doctor of Philosophy levels. His research interests are in artificial intelligence applications such as in the fields of acoustics, modeling, and control. He was granted a Charted Engineer status by Engineering Council, United Kingdom, in 2004.